REVIEW

English symbol — Greek letters

Statistic	Parameter	
$\bar{x}$	μ	means
s	σ	std. dev.
$\hat{p}$	p rho	prop.

Test Statistic:
Generic Form:

$$\frac{\text{statistic} - \text{parameter}}{\text{std. dev. of statistic}}$$

(SE)

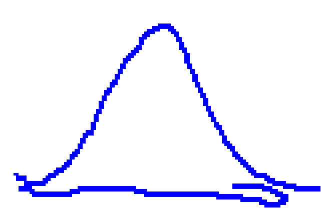
Confidence Interval:
Generic Form:

$$\text{statistic} \pm \begin{pmatrix}\text{critical}\\ \text{value}\end{pmatrix}\begin{pmatrix}\text{std. dev}\\ \text{of stat.}\end{pmatrix}$$

(SE)

What are assumptions used for? To check for what?

- check if conditions are met for inference



Remembering back to Ch. 1....

What's the formula for a Z-score when talking about averages?

$$Z = \frac{X - \mu}{\sigma}$$

I have an observation that has a Z-score of 1.4. What does this mean about the observation compared to its mean?

1.4 std. deviations above its mean

40 year old females have heights that are N(65, 2.5). What percent of females are 66" or above? Don't forget notation!

μ σ $n = 20$

$$P(X > 66") = \text{normcdf}(66, E99, 65, \frac{2.5}{\sqrt{20}}) = 0.3446$$

what ht has 10% of females taller than it?

$$\text{invnorm}(0.90, 65, 2.5) = 68.204"$$

Try the problems on the next page as a review

1) a) $z=2$, $z=-3$

b) $x=1.8925$, $x=1.72$

2) a) $z=1.375$

b) 694

3) a) $P(x>16)=0.9938$

$P(x<L)=0.10$

$L=16.05$

d) 

? ?

16.034 16.168

- In the problems above, you were always given μ and σ and not n.

unknown

- However in statistics, we always make conclusions from samples.
- When we have a sample, we don't have μ and σ anymore, we have $\bar{x}$ and s.
- So if a population has a distribution: $N(\mu, \sigma)$

  then the sample has a distribution: $N(\bar{x}, s)$

- Where did we see problems like this before?

Ch.5 - sample means

$\bar{x}$

Example of what we've seen before:

Cola bottles are supposed to be filled with 300mL of soda. However, bottles are usually normally distributed with an average of 298 mL and a standard deviation of 3mL.

1- What is the distribution of one bottle?

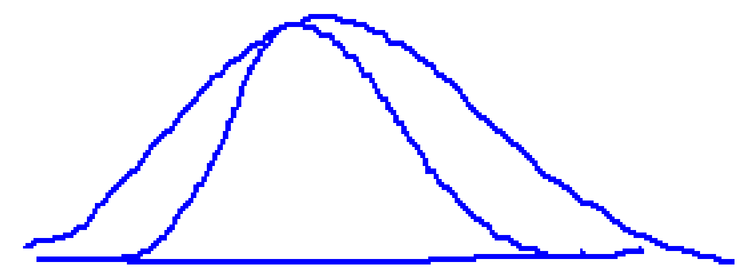
$N(298, 3)$

2- What is the probability of selecting one bottle that is less than 295 mL?

$P(X < 295) = \text{normcdf}(-E99, 295, 298, 3) = 0.1587$

3- What is the distribution of a 6-pack?

$n = 6$ $\quad N\left(298, \frac{3}{\sqrt{6}}\right)$



4- What is the probability of selecting a 6-pack that has an **average amount of cola less than 295 mL?**

$P(\bar{x} < 295) = \text{normcdf}\left(-E99, 295, 298, \frac{3}{\sqrt{6}}\right) = 0.00715$

Ch. 7: The T-Test

- This chapter, we are looking to make conclusions about the unknown parameter …

  μ (population mean)

- μ is estimated by ___ $\bar{x}$ ___
- Since we don't know μ, then we don't know ___ σ ___ either. Why?

  μ is in formula for σ

- σ is estimated by ___ s ___
- so, if a population has a distribution of N(μ,σ), we can estimate that by: $N(\bar{x}, s)$
- But what if we take a sample of size *n*? How does that affect our estimate of the population?

  $N\left(\bar{x}, \frac{s}{\sqrt{n}}\right)$

- $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ = std. dev. of $\bar{x}$ = std. dev. of statistic

- But we don't know σ. We estimate σ with ___ s ___

- Thus, we will estimate $\sigma/\sqrt{n}$ with: $s/\sqrt{n}$ = std. dev of statistic = $SE_{\bar{x}}$

- Once we do this estimation, we can't use ___ z-distribution ___. Why?

z distr. is based on μ & σ (pop.)

z-scores

normcdf

$z = \dfrac{x-\mu}{\sigma}$

- Instead, we will use what is called the ___ t-distribution ___.

samples

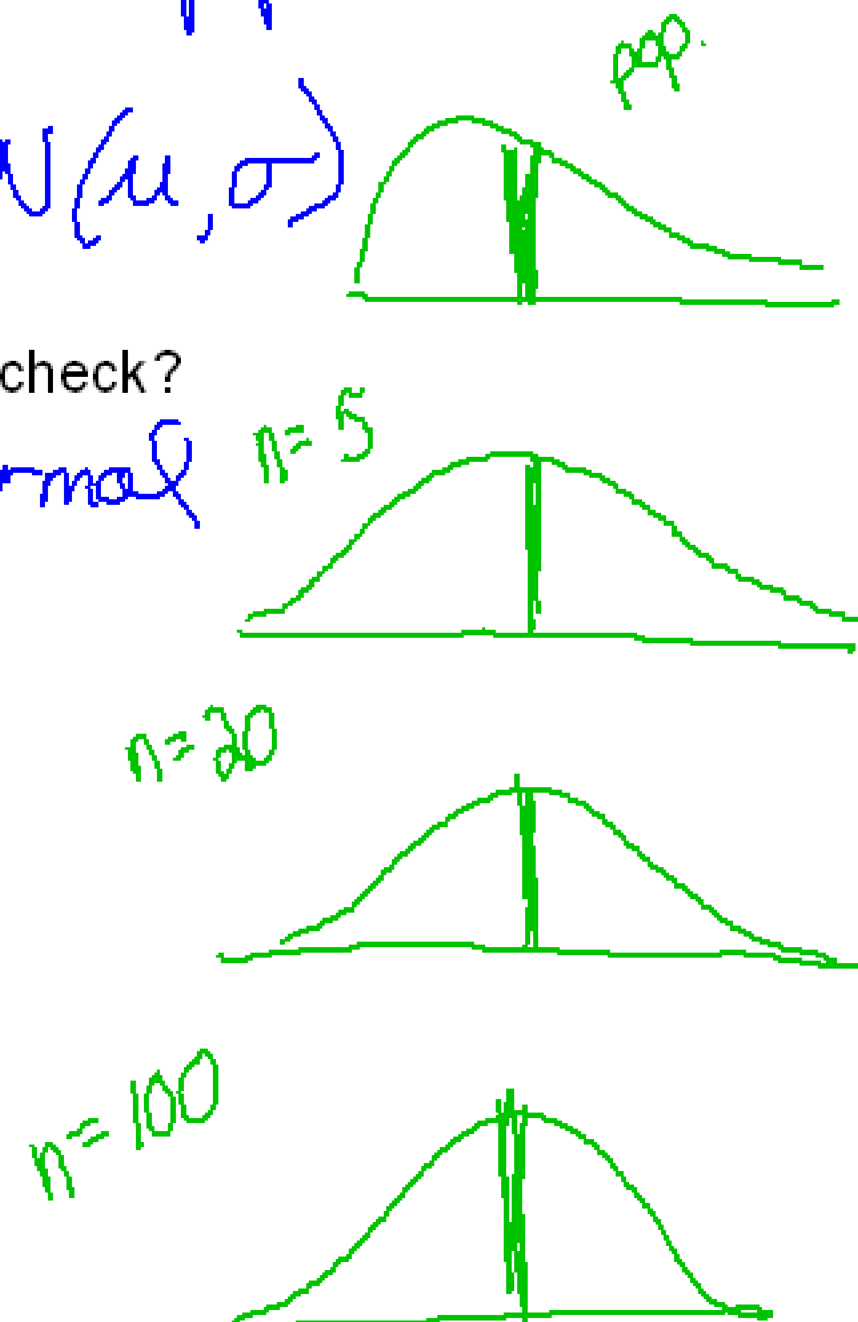
The T-distribution:

$s/\sqrt{n}$

- Still… taking an SRS of size n from pop.
- Still… from normal pop. with $N(\mu, \sigma)$
  - If we don't have a normal population, what do we check?

    $n \geq 30 \Rightarrow$ sampling distr ~ Normal
- There is … a different t-distribution for every sample size.

know n



- Similarities to Standard Normal Curve (used for z scores):
  - symmetric, bell-shaped
  - centered @ 0
- Big difference from Standard Normal Curve
  - t distribution is more spread out (wider)
  - Why?

$\underbrace{S \text{ more variable}}_{\text{changes}}$ than σ (doesn't change)

- How does the sample size affect the t-distribution?

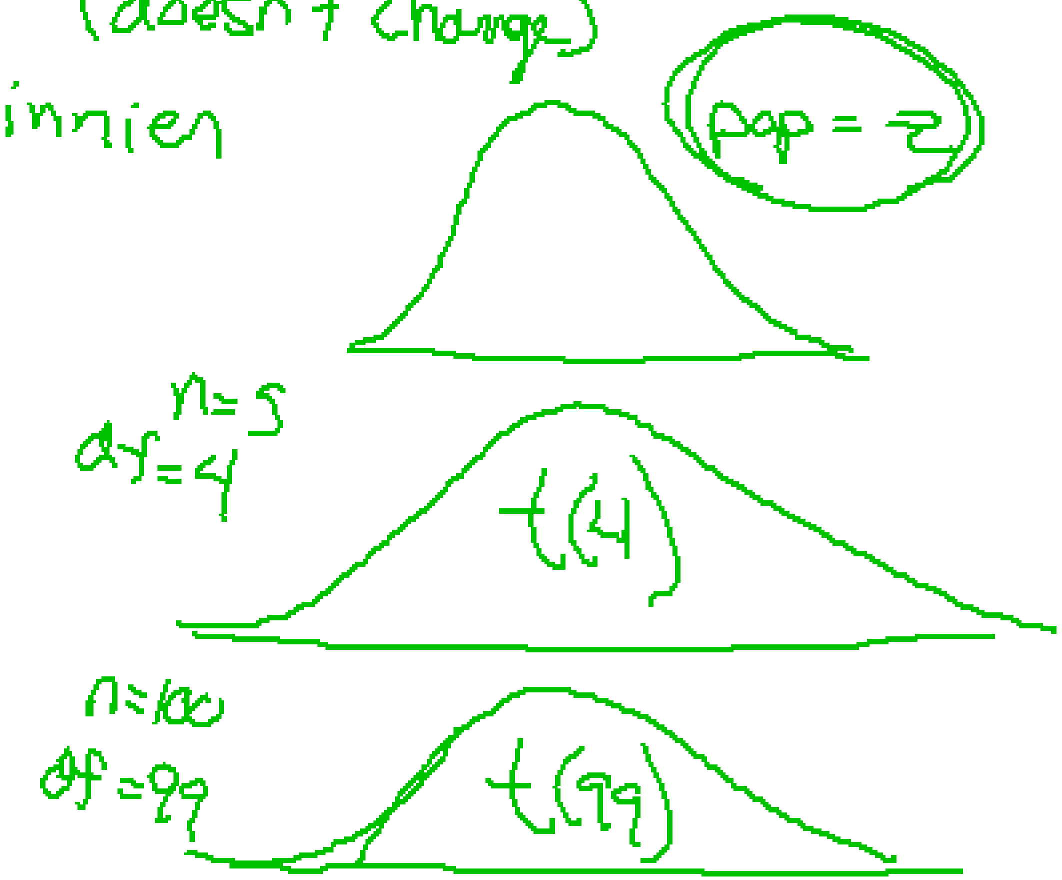
in general, as $n \uparrow$, t = skinnier

  - Degrees of freedom = $n-1 = k$
  - As the degrees of freedom increase...

(as $n \uparrow$) $t(k)$

t distribution becomes closer to z-distrib (pop)

$t(\infty) = z$ distr.



The One Sample T-Test

Same steps:

1. Assumptions
2. Hypotheses
3. Test Stat.
4. P-value
5. Conclusion

Hypotheses:

- Similar to proportion tests:

H_0: parameter $=$ #
H_a: parameter $<$ / $>$ / $\neq$ #

H_0: $\mu =$ #
H_a: $\mu >$ / $<$ / $\neq$ #

- Except they are about…

means

Test Statistic:

MEANS

- Formula: Generic

statistic − parameter / Std. deviation of statistic

$$\frac{\text{statistic} - \text{parameter}}{\text{Std. deviation of statistic}}$$

Specific:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Ho

P-Value:

- Notation: $P(t \gtrless \underline{\text{test stat}} \mid df = \quad) =$

- Calculator Use:

tcdf(LB, UB, df)

Conclusion:

- Same 2 sentences, except... abt means

- Reject/Fail to reject b/c p-value...
- We have suff. evid that the mean of context is ≥ ≤ ≠ # units.

Add to your notes...

Example:
It has been claimed that the average GPA for AP students is 3.3. We dont believe this claim- we think that since more and more students are taking AP classes and not taking the AP test, the average GPA has gone down. We take a sample of 120 AP students and ask them their GPA. We find that this sample has an average GPA of 3.1 and a standard deviation of 0.15. Test this claim at the 0.5 significance level.

$H_0: \mu = 3.3$ $\quad n = 120 \quad s = 0.15$

$H_a: \mu < 3.3$ $\quad \bar{x} = 3.1 \quad \alpha = 0.05$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = -14.606$$

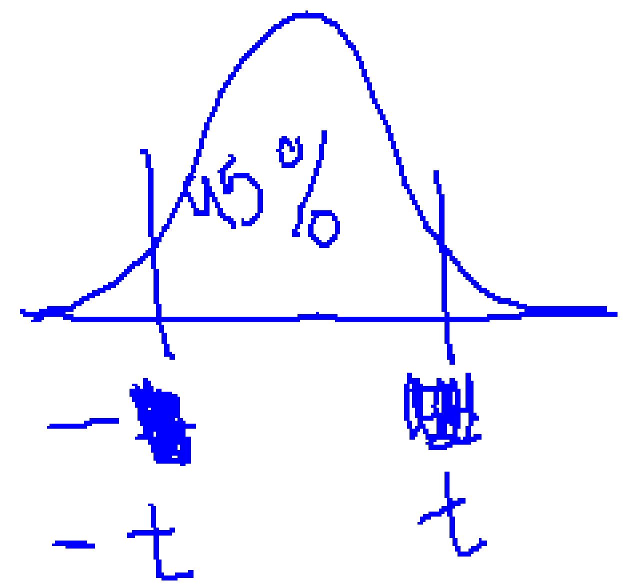
$$P(t < -14.606 \mid df = 119) = 0$$

- We reject Ho in favor of Ha b/c p-val of $0 < \alpha = 0.05$.
- Suff. evid that the avg. GPA of AP students is less than 3.3 units.

t Confidence Interval MEANS

Formula: generic statistic ± (critical value)(std. dev. of stat.)

Specific: $\bar{x} \pm t^*\left(\frac{s}{\sqrt{n}}\right) = (a, b)$



- t^* = ± t that has ___ % of data btw

- How do we find t^*?

  - table B
  - closest df

Conclusion:

- Same as with z-interval

We are ____% conf. that the mean of ______

is btw ______ and ______ units.

Assumptions for t-test and t-interval

check

1. SRS

   1. circled/assumed

2. normal pop or $n \geq 30$

   2. circled or $n = 100 \geq 30$

np

$n(1-p) \geq 10$

Some vocab for this chapter…

Robustness…

- when we can still use an inference procedure (test or int) even if assumptions aren't met

*none in this class

- * like resistant, for inference

Add to your notes...

Example:
Since the housing market has changed so drastically over the past few years, we don't know what the average price for homes in Bucks County is anymore. We take a sample of 90 recently sold homes and look at their selling prices. We find that our sample has an average price of $313,562 and a standard deviation of $70,945. Using a 96% confidence interval, find an estimate for the average price of homes in Bucks county.

$n = 90$

$\bar{x} = \$313{,}562$

$s = \$70{,}945$

Conf = 96%

Assump

1. SRS
2. normal pop or $n \geq 30$

Check

1. assume
2. $n = 90 \geq 30$ ✓

$\bar{x} \pm t^* \, s/\sqrt{n}$

$= (297{,}975, \; 329{,}149)$

We are 96% conf. that the avg. price of homes in Bucks is btw $297,975 and $329,149.

	one sample Z test	one sample t test
parameter we are estimating	p	μ
pop. std. deviation	$\sqrt{\frac{p(1-p)}{n}}$	σ
sample std. deviation	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$s/\sqrt{n} = SE$
Test Statistic	$z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$	$t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$
Confidence Interval	$\hat{p} \pm z^{*}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$\bar{x} \pm t^{*}\left(s/\sqrt{n}\right)$

Distribution: Name	Std. normal distr.	t-distribution
Distribution: Center	0	0
Distribution: Spread	1	depend on n (df)
Other	—	$df = n-1$
Assumptions	1. SRS 2. np, $n(1-p) \geq 10$ 3. pop $\geq 10 \cdot n$	1. SRS 2. normal pop or $n \geq 30$

z	t
1 prop z test z int	T-test T-Interval

Try the worksheet on the next page on your own